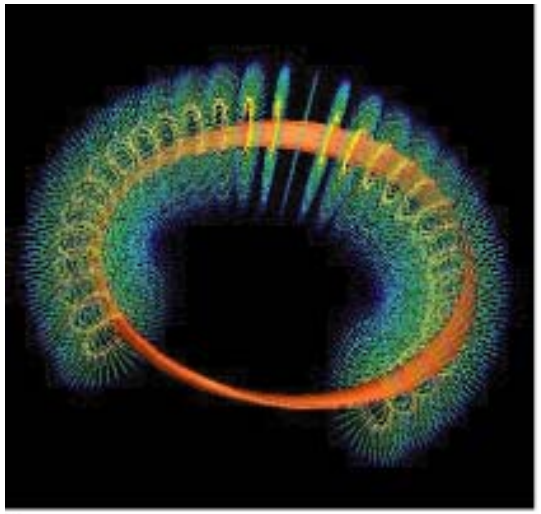
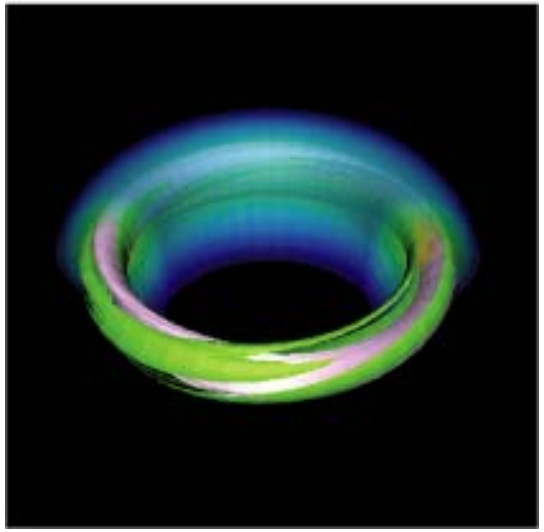




THE NATIONAL FUSION COLLABORATORY:

Grid Computing for Simulation and Experiments

Presented by
Martin Greenwald
APS – May 3, 2004

http://www.fusiongrid.org

Nature of Fusion Research Drives Requirements for Computing and Networking

- Experiments
  - Characterized by near real-time interactions of large, geographically extended teams
  - Faster between-pulse analysis translates directly to productivity
  - Barriers to use of powerful analysis tools can be significant
- Theory and Computation
  - Simulations producing very large data sets (GB $\Rightarrow$ TB $\Rightarrow$ PB)
  - Interactive visualization and analysis present a severe challenge for computing and networking
  - Increased code sharing and collaborative development

FusionGRID
www.fusiongrid.org

Collaboratory's Goal is to Advance Scientific Understanding & Innovation in Fusion Research

- Enable more efficient use of experimental facilities through more powerful between pulse data analysis
- Allowing transparent access to analysis and simulation codes, data, and visualization tools, resulting in more researchers having better access to more resources
- Enable more effective integration of experiment, theory, & modeling
- Facilitate multi–institution collaborations
- Create a standard tool set for data access, security, and visualization allowing researchers to build these into their own applications

FusionGRID
www.fusiongrid.org

Who is Working on the Collaboratory?

- **Fusion Labs**
  - General Atomics
  - MIT – Plasma Science & Fusion Center
  - Princeton Plasma Physics Laboratory

- **Computer Science Labs**
  - Argonne National Laboratory
  - Lawrence Berkeley National Laboratory
  - Princeton University
  - University of Utah

(Funded by OASCR SciDAC program)

FusionGRID
www.fusiongrid.org

We Are Not Focusing on "Traditional" Grid Applications – Cycle Scavenging and Dynamically Configured Server Farms

- Traditional Computational Grids – Arrays of heterogeneous servers
- Machines can arrive and leave
- Adaptive discovery – problems find resources
- Workload balancing and cycle scavenging
- Bandwidth diversity – not all machines are well connected

- **This model is not especially suited to fusion computation**
- **We are aiming to move high-performance distributed computing out onto the wide-area network**

FusionGRID
www.fusiongrid.org

Putting Distributed Computing Applications out on the Wide Area Network Presents Significant Challenges

- Crosses administrative boundaries
- Increased concerns and complexity for security model – (authentication & authorization)
- Resources not owned by a single project or program
- Distributed control of resources **by owners** is essential
- Needs for end-to-end application performance and problem resolution
  - Resource monitoring, management and troubleshooting are not straightforward
  - Higher latency challenges network throughput, interactivity
- People are not in one place for easy communication

FusionGRID
www.fusiongrid.org

Vision for the Fusion Collaboratory

- Data, Codes, and other resources should be thought of as network accessible services.
- Shared security infrastructure – with distributed authorization and resource management
- Collaborative nature of research requires shared visualization applications and widely deployed A/V technologies

- We are not focused on CPU cycle scavenging or “distributed” supercomputing (typical GRID justifications)

Optimize the most expensive resource - people’s time

FusionGRID
www.fusiongrid.org

Vision – Resources as Services

- **Resources = Computers, Codes, Data, Analysis Routines, Visualization tools, Experimental Operations**
- Access is stressed rather than portability
- Users are shielded from implementation details.
- Transparency and ease-of-use are crucial elements
- Shared toolset enables collaboration between sites and across sub-disciplines.
- Knowledge of relevant physics is still required of course.

FusionGRID
www.fusiongrid.org

Vision – Security Infrastructure

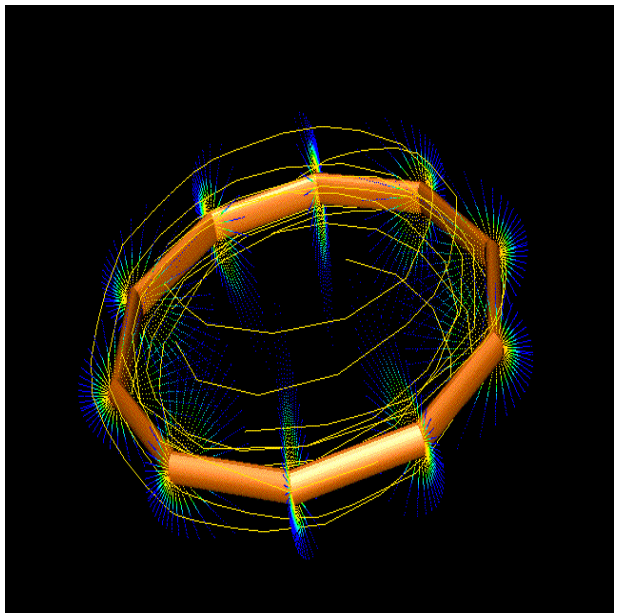
- Strong authentication identifies users (currently based on x.509 certificates from DOE Science Grid).

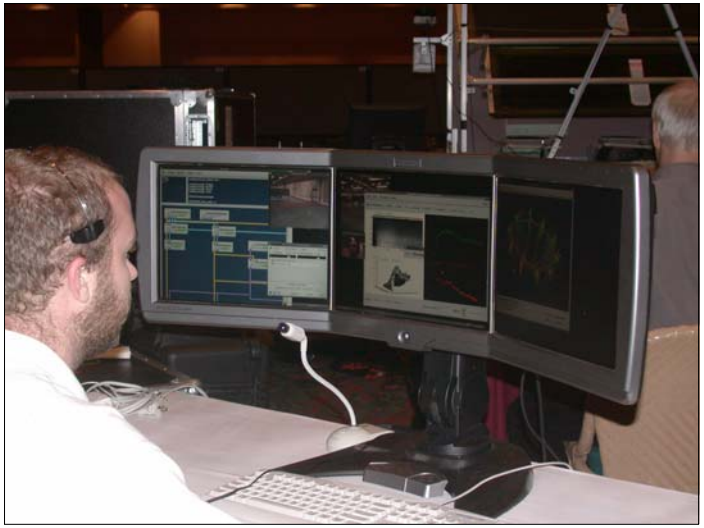
- Distributed authorization allows stakeholders to control their own resources.
  - Facility owners can protect computers, data and experiments
  - Code developers can control intellectual property
  - Fair use of shared resources can be demonstrated and controlled.

FusionGRID
www.fusiongrid.org

Vision – Visualization and A/V Tools

- Maximum interactivity for visualization of very large data sets



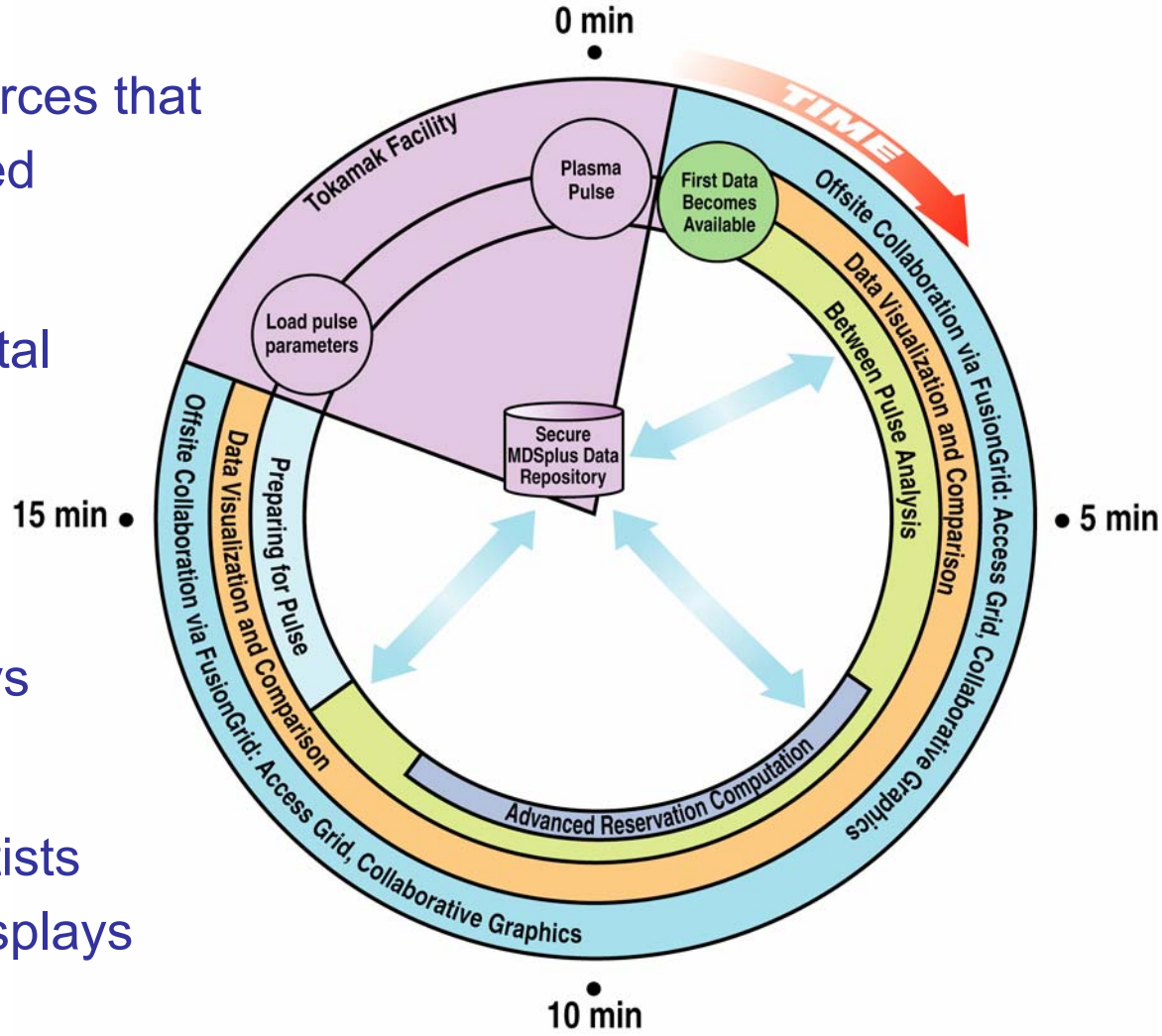


- Use of extended tool sets for remote participation
  - Flexible audio and video links
  - Shared applications

FusionGRID
www.fusiongrid.org

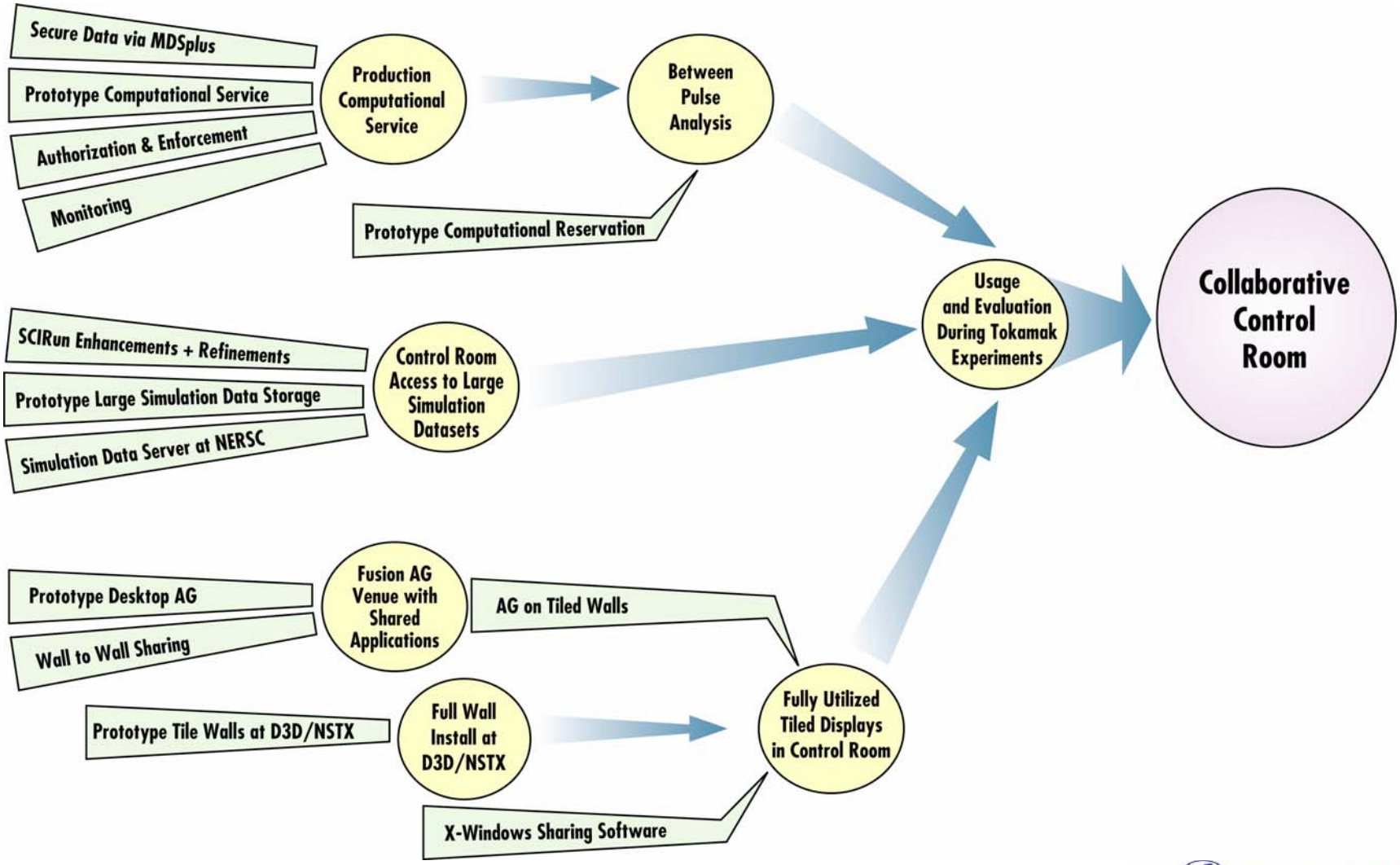
Collaboration is Fundamental to Advancing Fusion Science

- Secure computational resources that can be scheduled as required
- Rapidly compare experimental data to simulation results
- Share individual results with the group via shared displays
- Fully engaged remote scientists with audio, video, shared displays



FusionGRID
www.fusiongrid.org

For Experiments: Work Towards a Collaborative Control Room



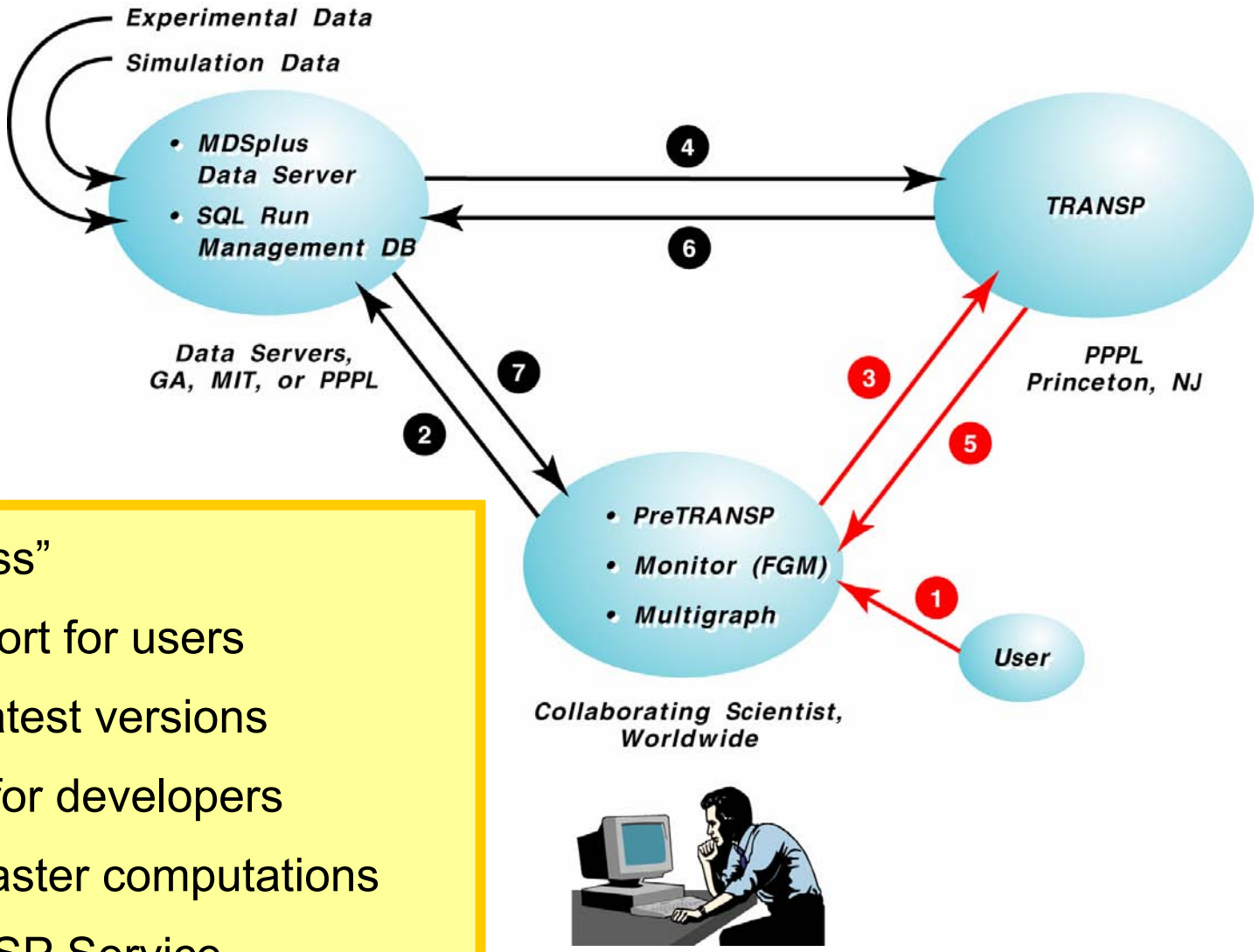
FusionGRID
www.fusiongrid.org

Secure Access to Data - MDSplus

- MDSplus: remote access based on client-server model
  - Used at more than 30 sites (robust) http://www.mdsplus.org
- "Service" rather than file oriented
- Hierarchical, self-descriptive, extensible, scalable, simple but powerful API
- MDSplus secured on FusionGrid via Globus GSI
  - Underlying technologies are X.509 certificates, OpenSSL
- Parallel network transfer via XIO – useful for high-latency links

FusionGRID
www.fusiongrid.org

TRANSP – First Grid Service Deployed



"This is a success"

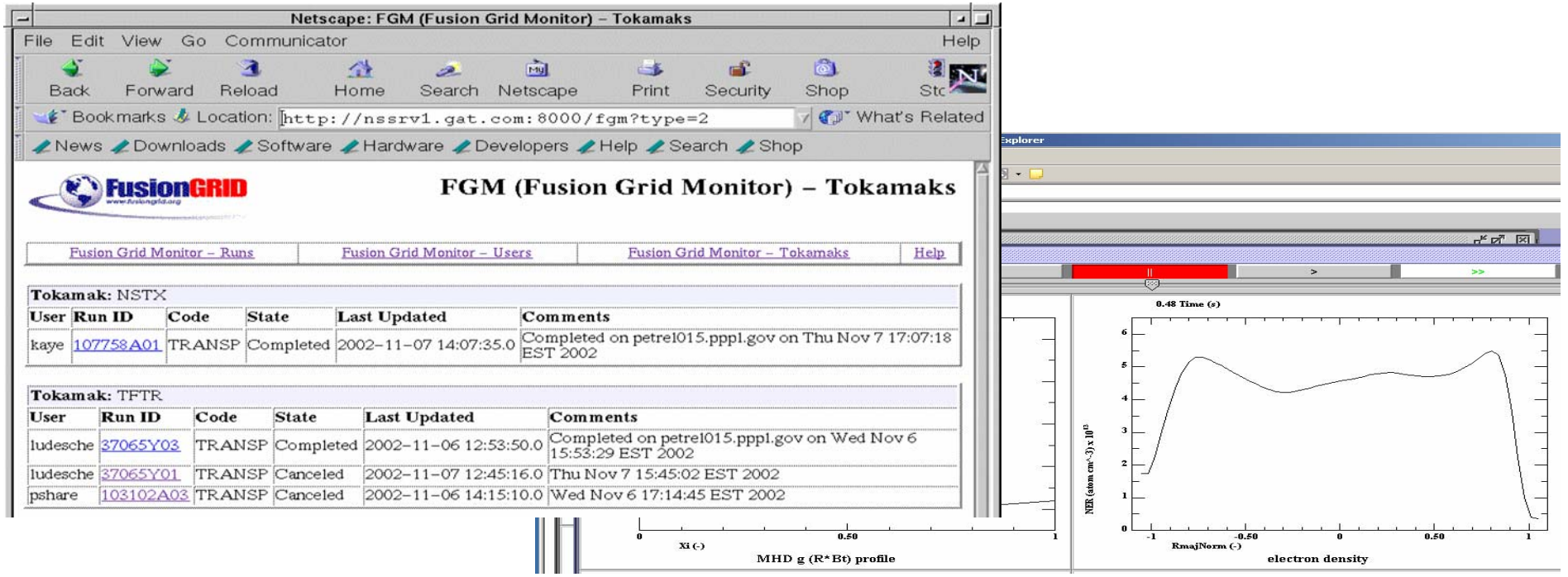
- Better support for users
- Users get latest versions
- Less effort for developers
- Access to faster computations

The U.S. TRANSP Service

- 1,800 runs, 9 tokamaks modelled

FusionGRID
www.fusiongrid.org

Fusion Grid Monitor: An Efficient Application Monitoring System for the Grid Environment

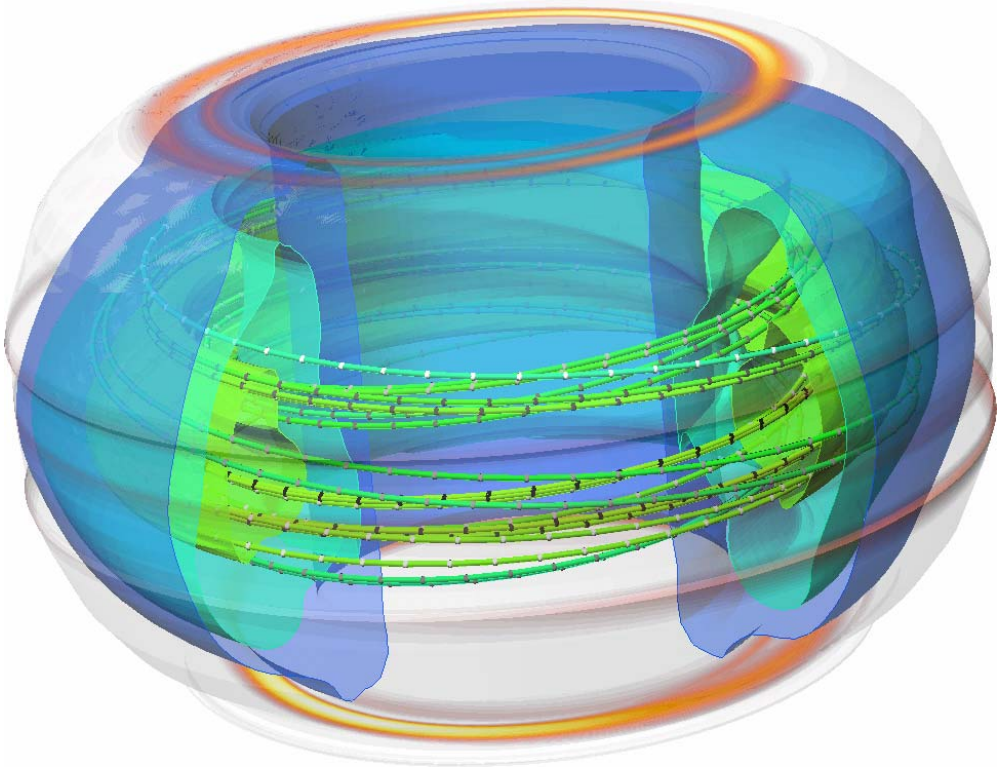


- Users track and monitor the state of applications on FusionGrid
  - Output dynamically via HTML, Built as Java Servlet (JDK2.1)
- Code maintenance notification
  - Users notified, queuing turned off, code rebuilt, queue restarted
- Results of simulation visualized during run

FusionGRID www.fusiongrid.org

SCIRUN: Visualize Complex Simulations

- Open source, multi-platform capable for a wide user base
- To facilitate quantitative comparison of simulations & experimental results



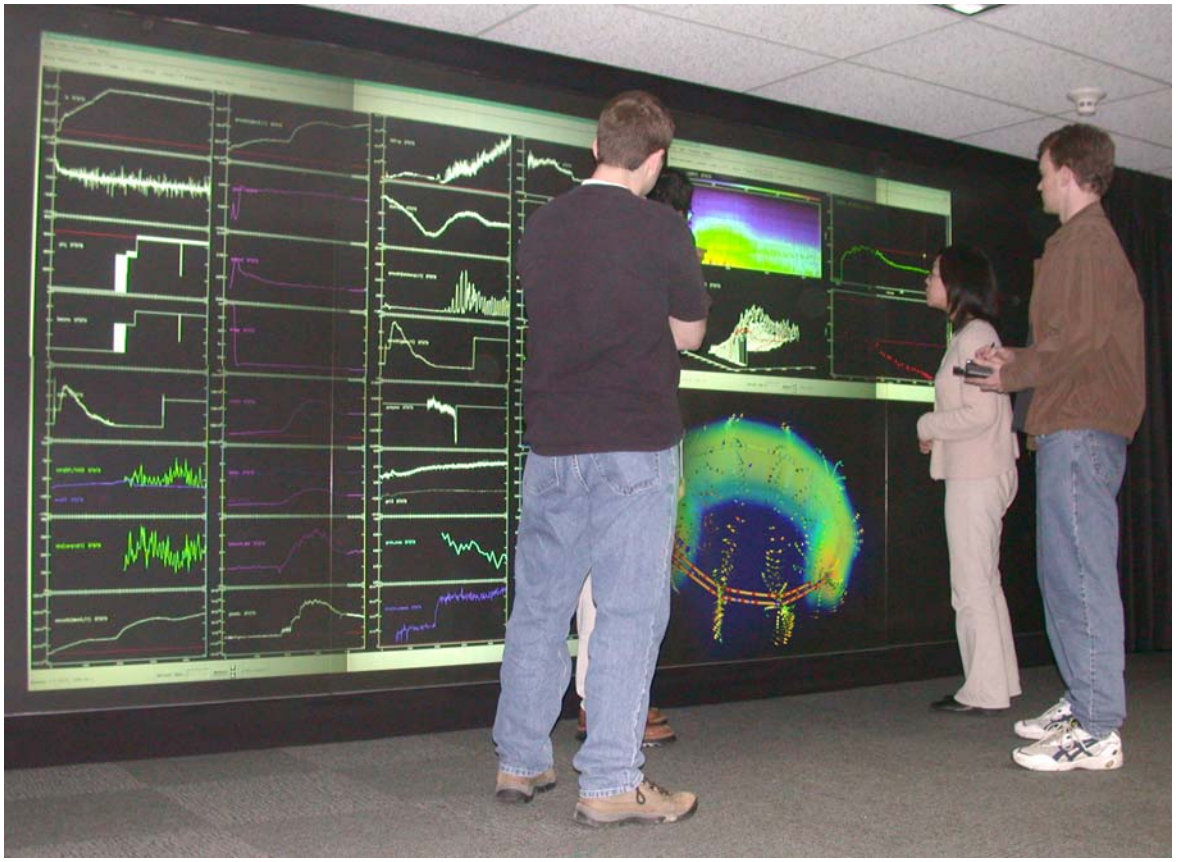
Raising the challenge of very large datasets

- Interactivity, latency
- MDSplus
- Storage method
- Data location
- Parallel I/O

FusionGRID
www.fusiongrid.org

Large Scale Displays and Shared Applications For Enhanced Collaborations

- Control room
  - Shared or remotely controlled displays
- Simulation Data
  - Access to ultra-high resolution images
  - Immersive environments



- Very well received by fusion scientists

  Fusion research funds used to purchase tiled walls for control rooms

FusionGRID
www.fusiongrid.org

Access Grid: High-End A/V tools

- Tested with off-site scientist to control room
  - Includes application sharing
  - Detailed data analysis discussion
- Feedback indicated the need for a greater control room presence for off-site scientists
- Should be useful for collaborations on computations

Personal Interface to the Grid (PIG) motivated by Fusion research

FusionGRID
www.fusiongrid.org

Summary

- Collaborative technology critical to the success of the FES program
  - Experimental: Fewer, larger machines in future (ITER)
  - Computation: Moving toward integrated simulation (FSP)
- The National Fusion Collaboratory Project is implementing and testing new collaborative technologies for fusion research
  - FusionGrid services being used daily to benefit FES research
- Clear vision forward to the collaborative control room
  - Concept encompasses most if not all collaborative FES needs

FusionGRID
www.fusiongrid.org